

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1/23/96	3. REPORT TYPE AND DATES COVERED Final Report, 10/1/94-9/30/95
4. TITLE AND SUBTITLE Structure and Sequence Coding in Proteins: Methods and Applications		5. FUNDING NUMBERS N00014-95-1-0002
6. AUTHOR(S) Shalom R. Rackovsky		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Dept. Of Biomathematical Sciences Mount Sinai School Of Medicine One Gustave L. Levy Place, Box 1023 New York, NY 10029		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office Of Naval Research 800 N. Quiney Street Arlington, VA 22217-5000		10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES		
12a. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Unlimited		12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 words) The characteristics of the inverse protein folding code, specifying the sequences which can be associated with a particular local fold, have been investigated. Methods development continues on the use of amino acid clustering for data compression. We are also investigating the relationship between protein structural symmetry and the distribution of amino acid properties along sequences.		

19960205 039

14. SUBJECT TERMS Inverse folding/amino acid properties/ amino acid clustering/protein structural symmetry			15. NUMBER OF PAGES 2
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRA UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)  
Prescribed by ANSI Std. Z39-18  
298-102

DTIC QUALITY INSURANCE

## Final Report

The principal thrust of the research funded by this award was continued elucidation of characteristics of the protein folding code. In work supported by another ONR award (N00014-91-J-1943) it was demonstrated that there exists a short-range folding code in proteins, and the characteristics of that code were investigated. In the present work, we have explored the possibility that there exists an inverse folding code in proteins. This is the code which determines which sequence fragments can be accommodated by a given local chain fold.

We have demonstrated<sup>1</sup> that an inverse folding code indeed exists, and that it has the same characteristics as the direct folding code (which addresses the structures permitted by a given sequence). In particular, it was shown that ~73% of four-residue structure fragments show a preference for particular distributions of sequences, and that the remainder do not. Rather, we suggest that these non-coding structure elements act as hot spots in which mutations are tolerated with no cost in *short-range* folding energy.

This is the first study to consider the inverse folding problem from the viewpoint of short-range interactions. All previous studies of this problem have considered the manner in which sequence substitution affects long-range interactions. Combination of our results with those of other workers, in a study of mutation frequencies in protein sequences, therefore has the potential to show the relative importance of short- and long-range interactions in determining mutation probabilities.

We are also continuing development work on methodology in this area. In recent computations, we have investigated the effect of amino acid clustering methods, useful for sequence data compression, on coding results.

A new area of interest initiated under funding from this award is the investigation of the connection between sequence properties and structural symmetry in proteins. Some protein structures show a high degree of three-dimensional symmetry, and physical intuition suggests that this structural property must be somehow reflected in the distribution of amino acid properties along the chain. We are developing Fourier methods to search for statistically significant periodicities of amino acid properties which may be linked to specific details of protein structure.

### Publications

1. S. Rackovsky, On the existence and implications of an inverse folding code in proteins. Proc. Nat. Acad. Sciences USA **92**, 6861 (1995).

### **Distribution List for Final Reports**

Attach a copy of the REPORT DOCUMENTATION PAGE (DD FORM 1473) to your final report as the first page and mail two copies (including the postcard labelled DTIC FORM 50 ) to:

Defense Technical Information Center  
Building 5, Cameron Station  
Alexandria, VA 22314

This allows other investigators to obtain copies of your report directly from DTIC. DTIC will fill out the postcard DTIC ACCESSION NOTICE (DTIC FORM 50) and return it to you with their number for your report. When you refer people to DTIC to get a copy of your report, give this number to expedite the request.

Mail one copy to each of the following and attach this very page to the back of your report - otherwise the folks below will think they have mistakenly received a copy meant for the Molecular Biology Program ):

- |   |  |
|---|--|
| (a) Dr. Michael Marron<br>ONR Code 1141<br>Molecular Biology Program<br>800 N. Quincy Street<br>Arlington, VA 22217-5000  | (e) Director<br>Chemical and Biological Sci Div<br>Army Research Office<br>P. O. Box 12211<br>Research Triangle Park, NC 27709 |
| (b) Administrative Grants Officer<br>ONR Resident Representative<br>(address varies - see copy of your<br>grant/contract) | (f) Life Sciences Directorate<br>Air Force Office of Scientific Res<br>Bolling Air Force Base<br>Washington, DC 20332          |
| (c) Director,<br>Applied Research Directorate<br>ONR Code 12<br>800 N. Quincy Street<br>Arlington, VA 22217-5000          | (g) Director<br>Naval Research Laboratory<br>Technical Information Div<br>Code 2627<br>Washington, DC 20375                    |
| (d) Director<br>Office of Naval Technology<br>Code 22<br>800 N. Quincy Street<br>Arlington, VA 22217-5000                 |  |